Is the world becoming a happier place?

An exploratory analysis of national happiness data
W200 Fall 23 Project 2 | Slides | Code Repo

Aaron Lin, Siddharth Manu, Stephanie Andrews

Introduction

This project has three main goals: to visualize global happiness levels by country and region, to analyze potential factors affecting national happiness, and to offer insights that could support global well-being.

Our main dataset is drawn from the World Happiness Report, which derives its scores from the Gallup World Poll—an extensive survey collecting responses from individuals in every country, with approximately one thousand respondents per country each year. The survey employs the Cantril Self-Anchoring Striving Scale, where respondents assess their subjective well-being by imagining a ladder with steps ranging from zero (representing the worst possible life) to ten (representing the best possible life). Gallup categorizes scores of 7 or higher as 'thriving,' scores between 4 and 7 as 'struggling,' and scores below 4 as 'suffering.' Notably, the lowest recorded life ladder score to date was 1.28, while the highest was 8.02.

For over a decade, the Sustainable Development Solutions Network has published the World Happiness Report, utilizing these life ladder scores and additional data on positive and negative emotions to formulate national happiness rankings. In our analysis, we examine raw life ladder scores to explore national happiness levels. Unlike the World Happiness Report, however, we do so without explicitly ranking countries, have scoped our analysis specifically to life ladder scores, and also try to explore potential contributing factors like climate and wealth inequality.

Research Question

What factors contribute to global happiness?

Sub-questions

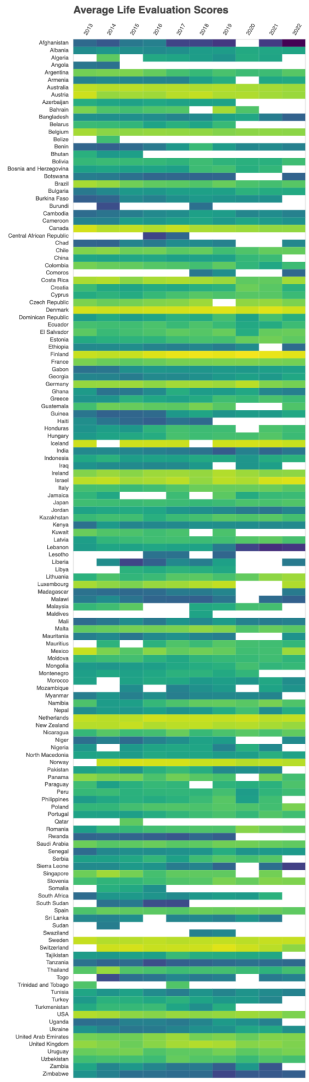
1. Is the world becoming a happier place?
2. Are wealthier countries happier?
3. Are countries with greater wealth inequality less happy?
4. Are countries with higher climate risk unhappier?
5. Is there a correlation between clean air and happiness?

Data Sources

- Primary: World Happiness Report Data
- Secondary: World Bank Population Data, World Inequality Database, Climate Risk Index, Environmental Performance Index, World Map GeoJson

Data Cleaning, Preprocessing and Assumptions

World Happiness Report Data



We examined two datasets from the World Happiness Report. One contained Cantril life ladder scores across all years for each participating country; the other had calculated measures for the most recent year (2022).

We pivoted, transposed, reindexed, and merged the core WHR datasets with supplemental datasets, detailed further below and in our repository's notebooks (eg `whr_data_prep.ipynb`).

Unfortunately, while both datasets were rich, they also were gappy. Not all countries participated in the annual surveys, and data was not available for some years among the countries that did participate.

This is a dataframe of WHR data after pivoting, that illustrates the number of values missing:

```
Index: 165 entries, Afghanistan to Zimbabwe
Data columns (total 18 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   2005    27 non-null     float64
 1   2006    89 non-null     float64
 2   2007    102 non-null    float64
 3   2008    110 non-null    float64
 4   2009    114 non-null    float64
 5   2010    124 non-null    float64
 6   2011    146 non-null    float64
 7   2012    141 non-null    float64
 8   2013    136 non-null    float64
 9   2014    144 non-null    float64
 10  2015    142 non-null    float64
 11  2016    141 non-null    float64
 12  2017    147 non-null    float64
 13  2018    141 non-null    float64
 14  2019    143 non-null    float64
 15  2020    116 non-null    float64
 16  2021    122 non-null    float64
 17  2022    114 non-null    float64
```

Our team discussed imputing missing values but concluded that doing so could be misleading, especially given some countries reported very seldomly (visible as white cells in the heatmap on left). So, for example, a score imputed for a country that reported once in the past decade would be less meaningful than for a country that only missed one reporting year and would be assigned an interpolated value. We further determined imputing was unnecessary at this stage of exploratory analysis.

World Inequality Database (WID)

Due to issues with exporting filtered data from the WID website, it was necessary to download the raw data compiled into 778 files comprising of 388 raw data files each representing a

different country or major region, 388 metadata files containing variable names and descriptions for each country, and 2 files containing indexes for continents and country names. To extract variables of interest, the file paths were obtained using the python os library and raw data file paths were extracted using regex through the python re library. Corresponding metadata files were identified using string concatenation and the pandas python library was used to load the csv files into data frames. From there, the metadata files were filtered to only variables of interest and the raw data file was filtered to only include the relevant percentiles of observed values. These data frames were joined together for each country and stored in a list that was concatenated to obtain one dataframe with columns of the desired variable names, variable values, percentiles, countries, and years.

Environmental Performance Index (EPI)
The EPI dataset consisted of individual files for each of the EPI factors per country. The data files corresponding to air quality factors were merged on the basis of country, code and ISO and using the weight ratios for each factor, air quality score was computed as a new column. There were no discrepancies related to merging as the dataset was complete. Some of the air quality scores were coming out to be negative, which is invalid and we discarded them during the cleaning process. Based on completeness of data and to avoid any complications related to the Covid 19 pandemic, we chose the data for the year 2018 to study the relationship between air quality and perceived happiness of a nation. Before merging, we had to update the country names for the United States as they differed between the WHR and the EPI datasets.

Climate Risk Index (CRI)
We cleaned up the column names of this dataset to remove special characters and updated the country names to match the world map GeoJSON for plotting the world map before we merged the two. In the step to study the relationship between climate risk and happiness, we had a lot of countries missing in the CRI and hence we had to drop a lot of countries.

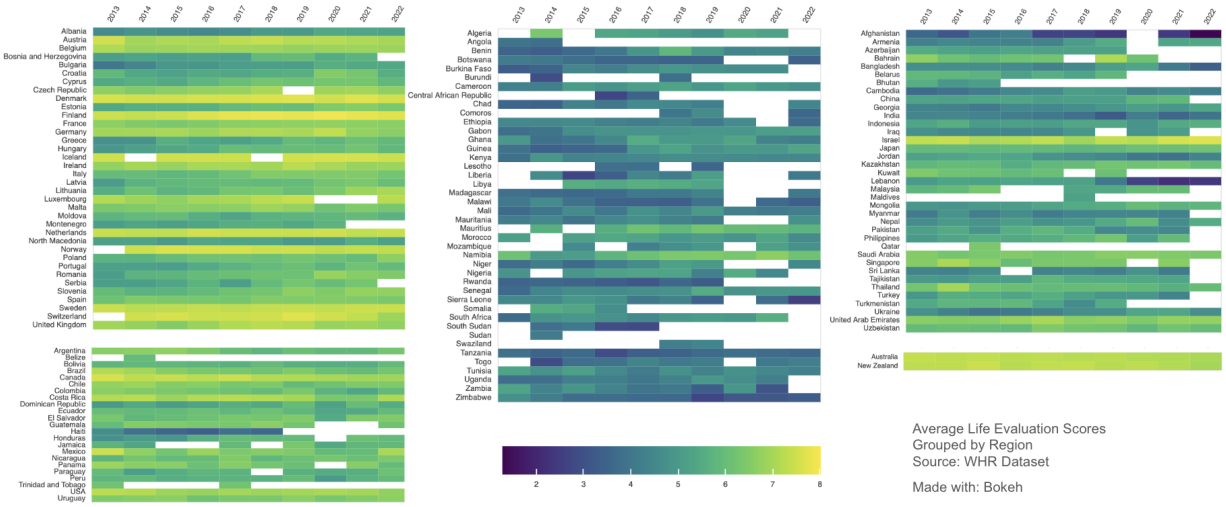
Analysis

Global Happiness Over Time

Data from the World Happiness Report was used in conjunction with World Bank national population data to study global happiness trends.

Q1: Is the world becoming a happier place?

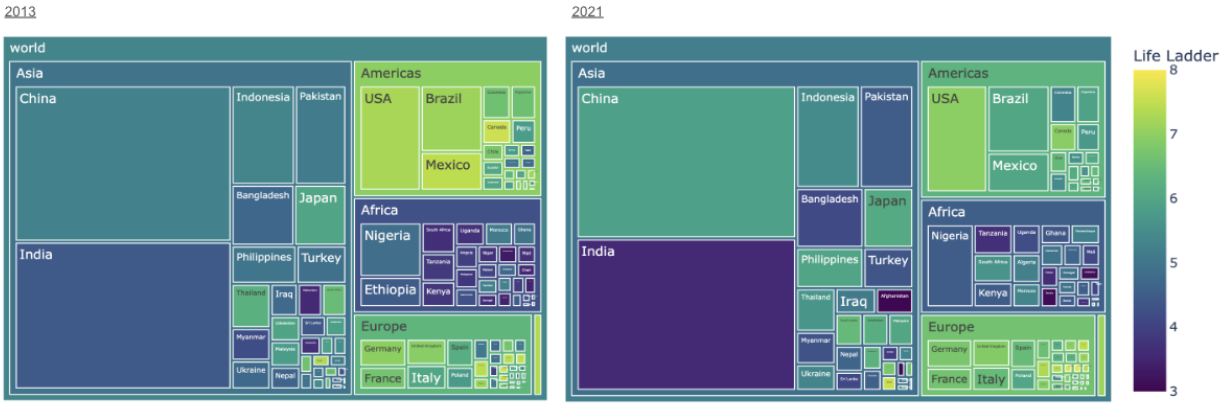
To begin addressing this question, we first generated heatmaps to visualize decade-long trends in life ladder scores, grouped by regions. Countries in Europe, Oceania, and the Americas consistently reported higher happiness scores relative to the rest of the world, while countries in Africa and some parts of Asia tended to report lower scores. For most countries, scores exhibited a downward trend over time, suggesting a decline in global happiness.

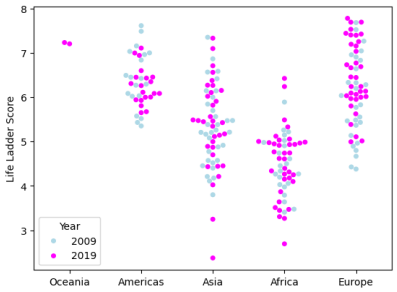


We then created several other visualizations comparing non-consecutive years, each consisting of two snapshots in time. Below is a set of scatter plots depicting the relationship between life ladder scores and population sizes in 2013 and 2022. A central cluster of countries with scores between 3.5 and 8 is evident in 2013. By 2022, there is a wider distribution, with several countries reporting notably lower life ladder scores than before.

We noticed China did not participate in the most recent survey (2022). So we next compared life ladder scores from 2013 against scores from 2021, as that was the most recent year we had data available for China.

The following two treemaps also illustrate life ladder scores in relation to population. Each block represents either a country or region, with size proportional to population and color scaled to the nation's average life ladder score for a given year. These treemaps suggest a decrease in national happiness for many countries, particularly for India and the more populous nations in the Americas. Several populous countries reported higher happiness scores in the same period, including China, the Philippines, and South Africa.





As our earlier comparisons (2022, 2021) may have been impacted by the pandemic, we wanted to also investigate changes in national happiness levels independent of this influence. To do this, we generated a swarm plot comparing scores from 2019, the year before the pandemic, with scores from a decade prior. Overall, there seems to be a global trend towards decreasing happiness in most regions.

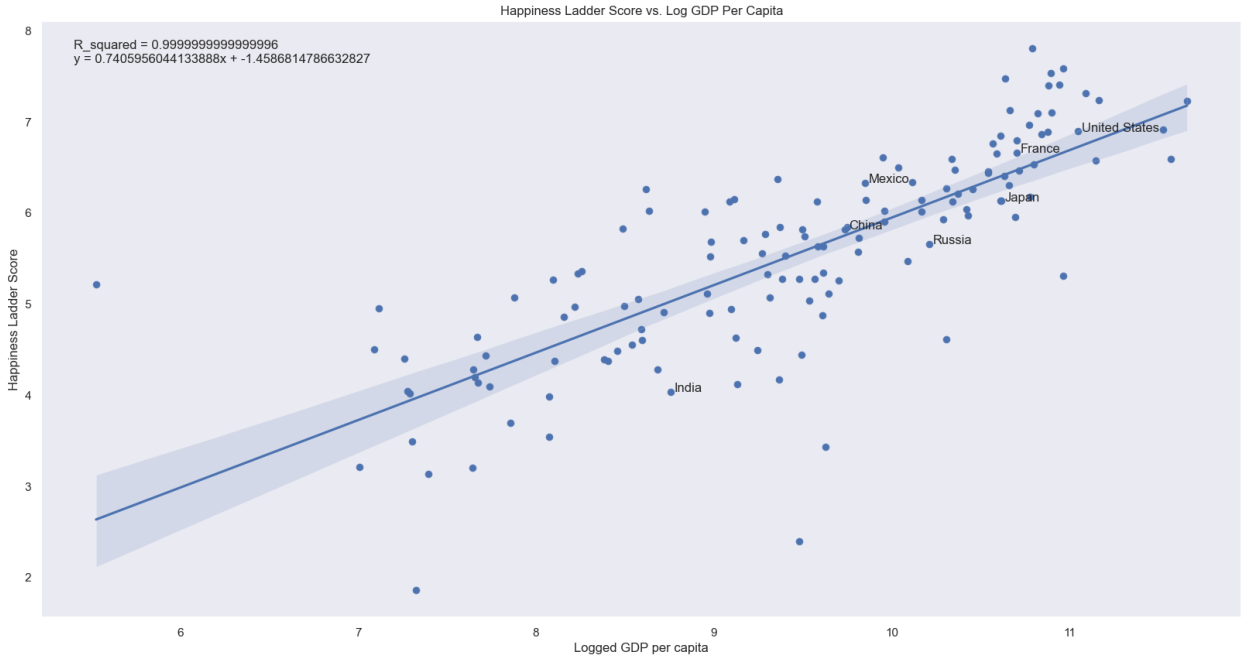
Our further analyses aim to explore factors that may have contributed to these fluctuations in global happiness, identifying potential areas where humanity can channel its efforts.

Economic Factors and Perceived Happiness

Data from the World Inequality Database was used in conjunction with World Happiness Report data to determine if there were any relationships between a country's economic status and the reported happiness of its citizens.

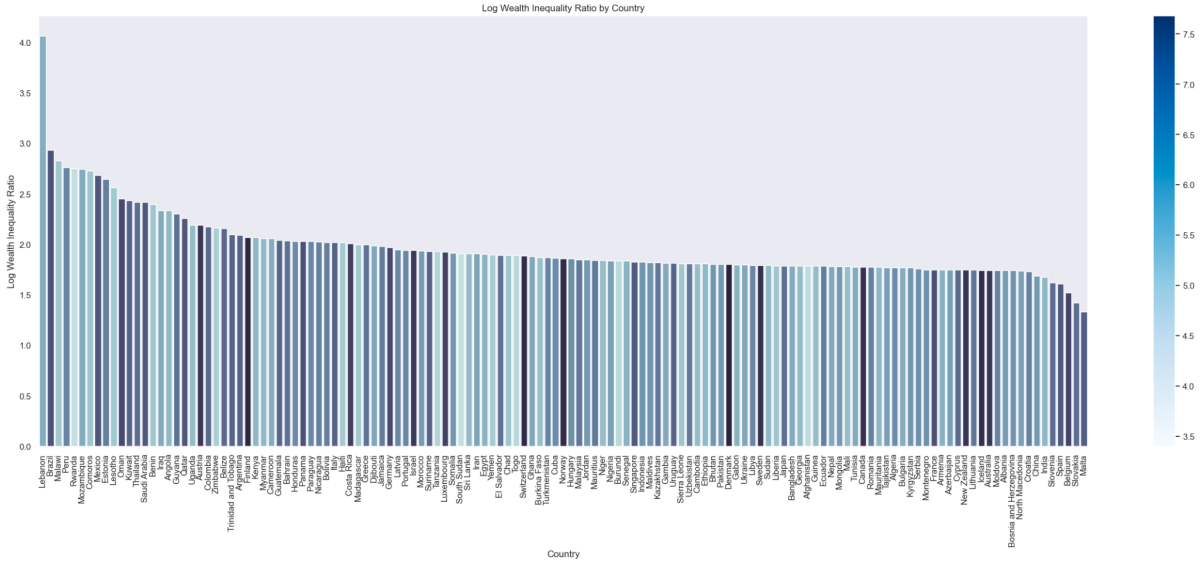
Q2: Are wealthier countries happier?

The world happiness report contains the logged Gross Domestic Product (GDP) per capita for recorded countries. The following plot shows the base 10 logarithm of the GDP per capita on the x-axis and the Happiness Ladder Score on the y-axis. The plot indicates that there is a tendency for the happiness ladder score to be higher for countries with higher logged GDP per capita values. A linear regression was performed to quantify the strength of this correlation which resulted in an r-squared value of over .99, indicating that there is a strong correlation between the logged GDP per capita and happiness as reported by the World Happiness Report.

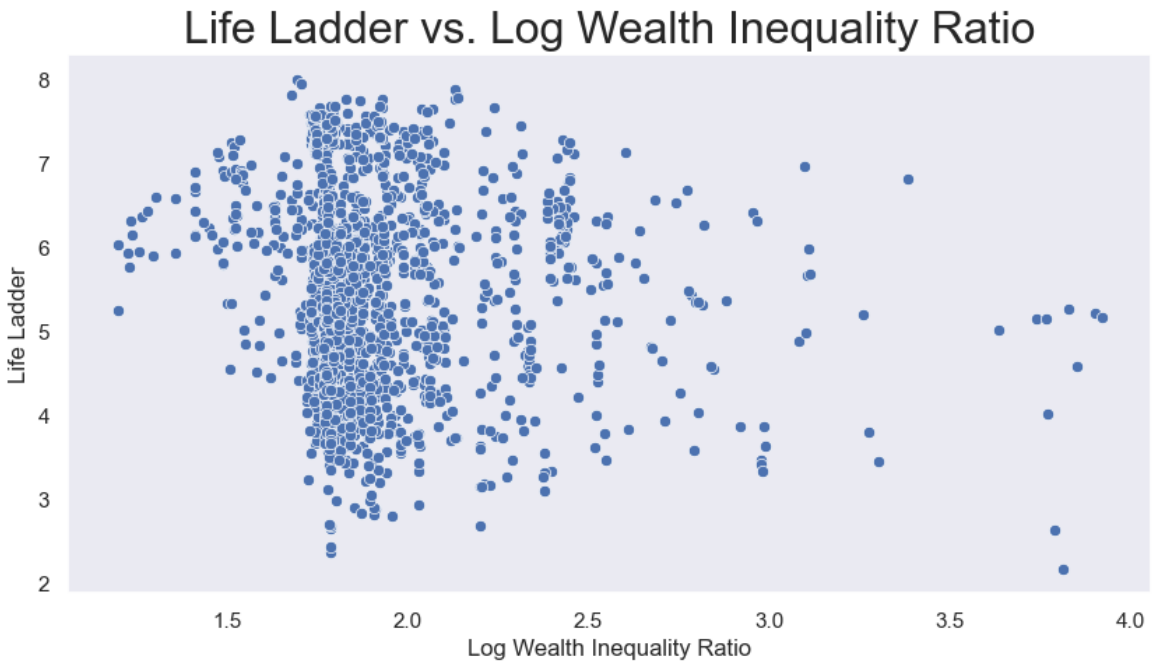


Q3: Are countries with greater wealth inequality less happy?

The second question our analysis sought to answer about the relationship between economics and happiness is whether or not Happiness values as reported by the WHR exhibit any relationships with wealth inequality. To answer this question, the World Inequality Database (WID) was used to extract the percentage share of private wealth owned by the top 10% and bottom 50% of the population. From these values, a ratio was calculated to represent how much more wealth the top 10% had than the bottom 50%. The median values for each country's time series data were plotted in a bar plot by country by descending order of Log Wealth Inequality Ratios with bar colors representing each country's corresponding happiness index values. Darker shades of blue indicate happier countries. This graph visualizes a relatively even distribution of happiness values across the spectrum, which suggests that there may not be a strong correlation between happiness and wealth inequality from the selected metrics.



The next visualization of this data is a scatter plot with the Happiness Values represented by the World Happiness Report's Life Ladder index on the Y-axis and the Log Wealth Inequality Ratio on the X-axis. This graph was prepared by plotting every country's data from 2005 to 2022. These cutoffs were selected from the World Happiness Report's earliest published values in 2005 and the latest reportest values from the World Inequality Database. One notable outlier was Lebanon in the year 2011, which had a ratio of over 90,000 represented by an x-value of 5 in this plot. This value was excluded due to the context of the year in the country's history; 2011 was a year of great turmoil in the country due to protests and the creation of a new government. In examining this scatter plot, we can see there is no discernable relationship between the selected variables though further analysis is necessary to create confidence in this claim.



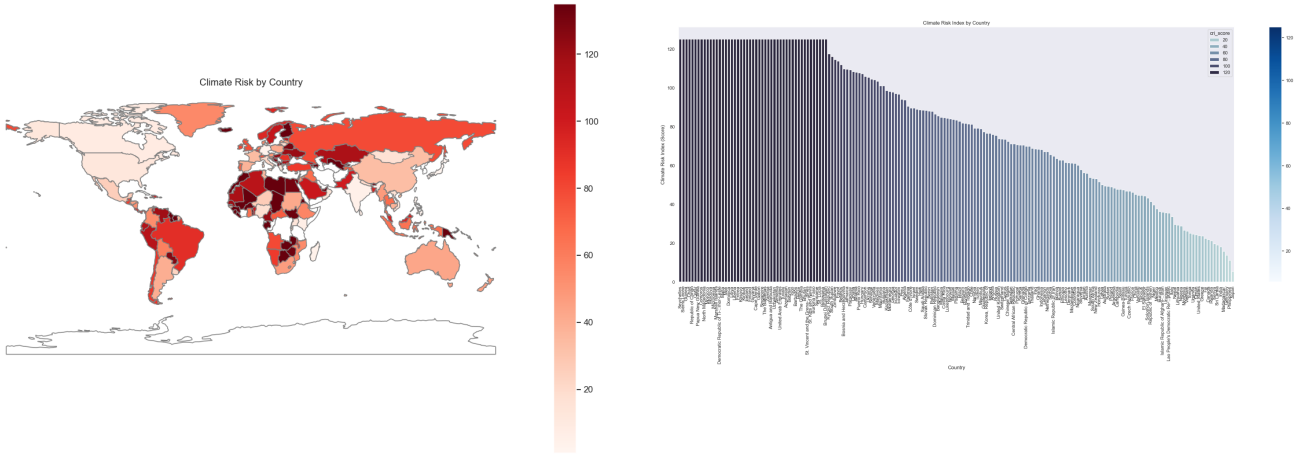
Environmental Factors and Perceived Happiness

Climate Risk Index and Environmental Performance Index datasets were used in conjunction with the World Happiness Report dataset to study the relationship between climate risk and happiness and environmental performance and happiness.

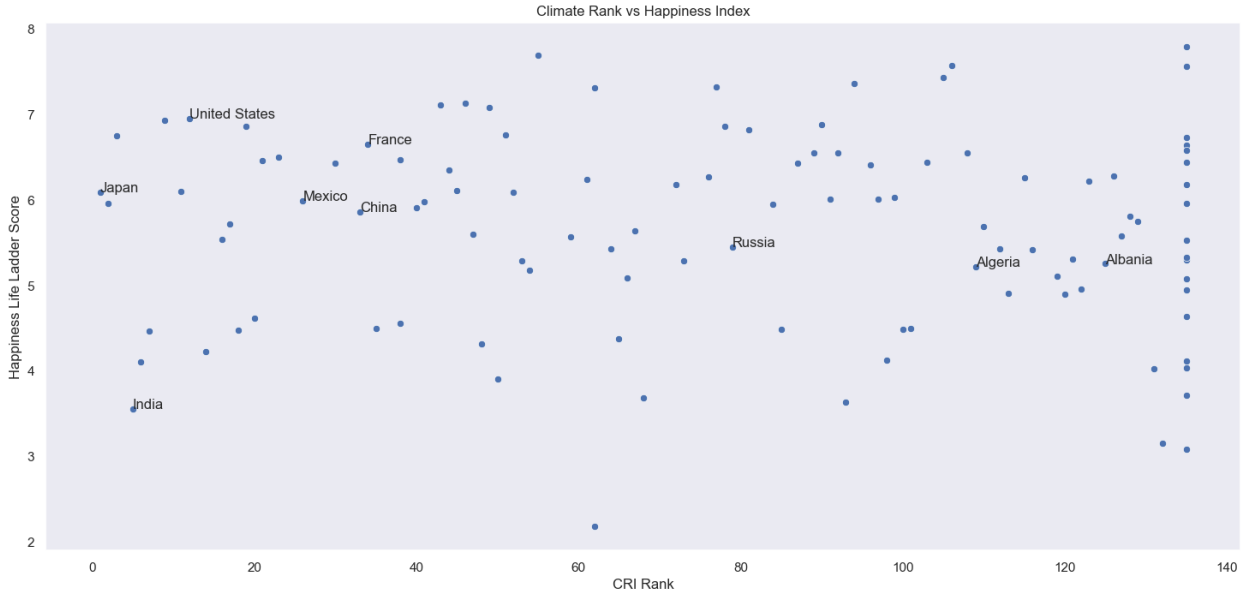
Q4: Are countries with higher climate risk unhappier?

We used the Climate Risk Index (CRI) to study the relationship between climate risk and happiness. The CRI is a tool that assesses and ranks countries based on their vulnerability to climate change impacts, considering factors such as extreme weather events. It helps identify nations at higher risk and informs strategies for climate resilience and adaptation.

In order to visualize the countries most vulnerable to climate risk, we plotted them on the world map using a color map to indicate the severity of the risk. This helped us see that countries in south-east Asia, parts of Europe, Africa, as well as South America face high risk. A bar plot of countries ranked by climate risk was also created to visualize the countries at most risk.



Next, in order to study the relationship between climate risk and happiness, we combined the data for the latest common year in the two datasets, 2021. We plotted CRI rank on the X-axis and Life Ladder score from the WHR on the Y-axis. We couldn't see any pattern here to show a strong correlation.



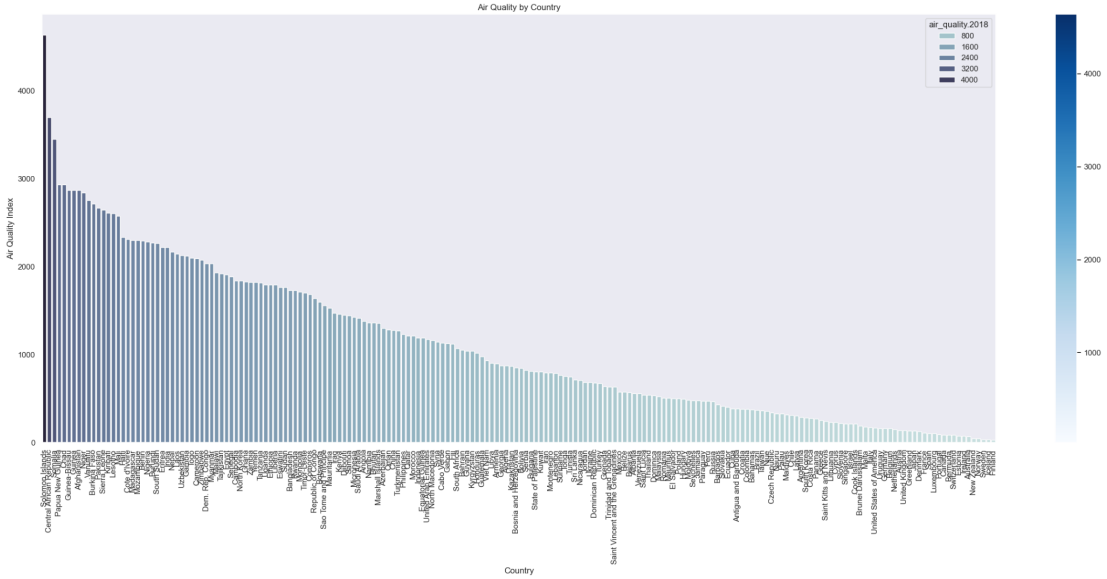
Q5: Is there a correlation between air quality and happiness?

The fifth and final question analyzed the correlation between air quality from the Environmental Performance Index dataset and happiness index from the WHR dataset. The air quality score was

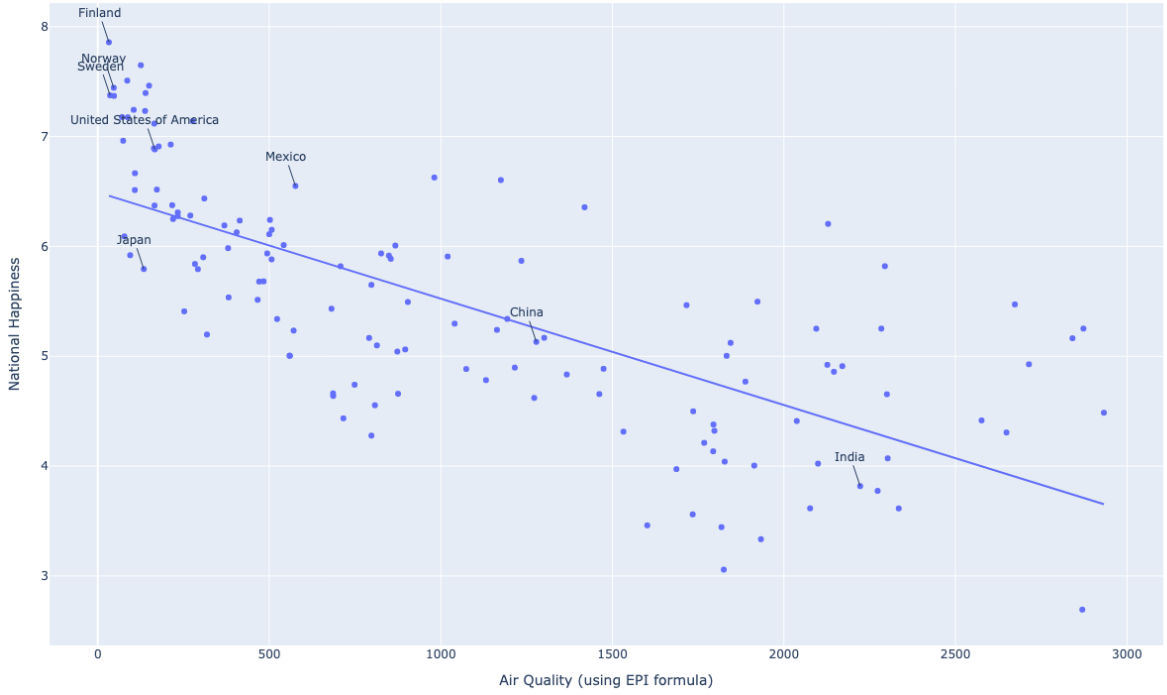
$PM_{2.5}$ Exposure	PMD	47%
Household Solid Fuels	HAD	38%
Ozone Exposure	OZD	5%
NOx Exposure	NOE	5%
SO2 Exposure	SOE	2%
CO Exposure	COE	2%
VOC Exposure	VOE	2%

calculated per country by merging together 6 datasets for each of the air polluting factors on the basis of country, and then aggregating them using the weight ratios prescribed in the EPI report for air quality.

The first visualization done was a bar chart on per-country air quality data sorted in descending order. The figure depicts that air quality greatly varies across countries.



Next, this data was combined with the happiness index per country from the WHR dataset and a scatter plot was generated along with a regression line based on the OLS technique. There was a strong correlation seen between the air quality and national happiness index.



Conclusion

Our analysis suggests that global happiness levels have been declining over time for many nations, even before the pandemic. We found strong correlations between a country's happiness index and factors such as GDP and air quality.

To deepen our comprehension of global happiness, our next steps include analyzing additional environmental factors such as clean water access, investigating correlations with conflict and terrorism, and assessing the influence of technological innovations on the world's well-being. Additionally, exploring social factors may reveal nuanced impacts on national happiness.

These insights can guide policymakers and global stakeholders in formulating targeted strategies to enhance well-being, focusing on areas like environmental policies, social programs, and innovation initiatives. This holistic approach ensures a comprehensive understanding and effective interventions to foster greater happiness in countries, ultimately, for a happier world!

References

- World Happiness Report
- World Happiness Report Data
- Gallup: “Understanding How Gallup Uses the Cantril Scale”
- World Bank Population Data
- Environmental Performance Index
- Climate Risk Index
- World Inequality Database
- World Map GeoJson